Report

Current Biology

Parallel functional reduction in the mitochondria of apicomplexan parasites

Highlights

- Single-cell transcriptomics of diverse apicomplexan parasites of invertebrates
- Reconstruction of the mitochondrial metabolic repertoire across apicomplexans
- Discovery of a range of reduced mitochondrion-related organelles in gregarines
- Multiple, parallel losses of electron transport and aerobic energy metabolism

Authors

Varsha Mathur, Kevin C. Wakeman, Patrick J. Keeling

Correspondence

varsha.mathur@botany.ubc.ca (V.M.), pkeeling@mail.ubc.ca (P.J.K.)

In brief

Mathur et al. show that gregarine apicomplexans have significantly reduced mitochondria: most lack complexes III and IV, and some lack the respiratory chain and TCA cycle entirely. Phylogenomics show that these reductions took place many times in parallel, resulting in a functional range from fully aerobic mitochondria to highly reduced mitosomes.

Mathur et al., 2021, Current Biology *31*, 2920–2928
July 12, 2021
https://doi.org/10.1016/j.cub.2021.04.028

CellPress

CellPress

Current Biology

Report

Parallel functional reduction in the mitochondria of apicomplexan parasites

Varsha Mathur,[1,3,*] Kevin C. Wakeman,[2] and Patrick J. Keeling[1,*]
[1]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[2]Institute for the Advancement of Higher Education, Hokkaido University, Sapporo 060-0810, Hokkaido, Japan
[3]Lead contact
*Correspondence: varsha.mathur@botany.ubc.ca (V.M.), pkeeling@mail.ubc.ca (P.J.K.)
https://doi.org/10.1016/j.cub.2021.04.028

SUMMARY

Gregarines are an early-diverging lineage of apicomplexan parasites that hold many clues into the origin and evolution of the group, a remarkable transition from free-living phototrophic algae into obligate parasites of animals.[1] Using single-cell transcriptomics targeting understudied lineages to complement available sequencing data, we characterized the mitochondrial metabolic repertoire across the tree of apicomplexans. In contrast to the large suite of proteins involved in aerobic respiration in well-studied parasites like *Toxoplasma* or *Plasmodium*,[2] we find that gregarine trophozoites have significantly reduced energy metabolism: most lack respiratory complexes III and IV, and some lack the electron transport chains (ETCs) and tricarboxylic acid (TCA) cycle entirely. Phylogenomic analyses show that these reductions took place several times in parallel, resulting in a functional range from fully aerobic organelles to extremely reduced "mitosomes" restricted to Fe-S cluster biosynthesis. The mitochondrial genome has also been lost repeatedly: in species with severe functional reduction simply by gene loss but in one species with a complete ETC by relocating *cox1* to the nuclear genome. Severe functional reduction of mitochondria is generally associated with structural reduction, resulting in small, nondescript mitochondrial-related organelles (MROs).[3] By contrast, gregarines retain distinctive mitochondria with tubular cristae, even the most functionally reduced cases that also lack genes associated with cristae formation. Overall, the parallel, severe reduction of gregarine mitochondria expands the diversity of organisms that contain MROs and further emphasizes the role of parallel transitions in apicomplexan evolution.

RESULTS AND DISCUSSION

The canonical view of mitochondria as oxygen-utilizing, ATP-generating organelles primarily derives from functional studies of aerobic model systems, such as animals, but nevertheless captures a central function across a wide spectrum of eukaryotic diversity. However, the universality of this picture was challenged with the discovery that anaerobic microbial eukaryotes that were thought to lack mitochondria[4] actually harbor relict, highly reduced, mitochondrion-related organelles (MROs).[5] The collective functional diversity of MROs has since been shown to cover a spectrum, ranging from aerobes or facultative aerobes lacking some complexes of the respiratory chain, to "hydrogenosomes" with anaerobic energy metabolism, to "mitosomes," which synthesize Fe-S clusters and have no role in energy metabolism.[6] This functional reduction can be linked to the loss of the mitochondrial genome[7–10] and is strongly correlated with structural simplification, making the most reduced MROs small and difficult to identify or recognize.[11,12] In only one known instance has the mitochondrion been lost altogether.[13]

In apicomplexan parasites, such as the well-studied and medically important pathogens, *Plasmodium* spp. and *Toxoplasma gondii*, the mitochondria retains canonical energy metabolism, including the tricarboxylic acid (TCA) cycle and ATP synthesis via an electron transport chain (ETC) similar to that of most other eukaryotes with the exception of complex I being replaced by an NADH dehydrogenase (NDH2).[14,15] However, their reliance on oxidative phosphorylation for ATP generation varies depending on life cycle stage, as demonstrated in *P. berghei*, where some components of the ETC are dispensable in the asexual blood stage but essential in the insect stage.[16,17] In contrast, the apicomplexan genus *Cryptosporidium* harbors MROs, which in the "intestinal-type" parasites, *C. parvum* and *C. hominis*, are functionally restricted to Fe-S cluster biosynthesis and in the "gastric-type" parasites, *C. muris* and *C. andersonii*, possess a broader spectrum of mitochondrial proteins, including the TCA cycle and a truncated ETC.[18] The relationship of *Cryptosporidium* to other apicomplexans remains contentious,[19] and until recently, the overall picture of apicomplexan phylogeny was restricted by the absence of data from many of the early-diverging groups, particularly the parasites of invertebrates for which no culture systems exist.[20] Single-cell transcriptomics has begun to fill these gaps and revealed the parallel evolution of many apicomplexan traits, including their plastid metabolism and even their parasitic lifestyle,[1,19,21] but their mitochondrial function has yet to be examined. To enable a detailed reconstruction of mitochondrial metabolism both between the major apicomplexan subgroups and within them, we sequenced

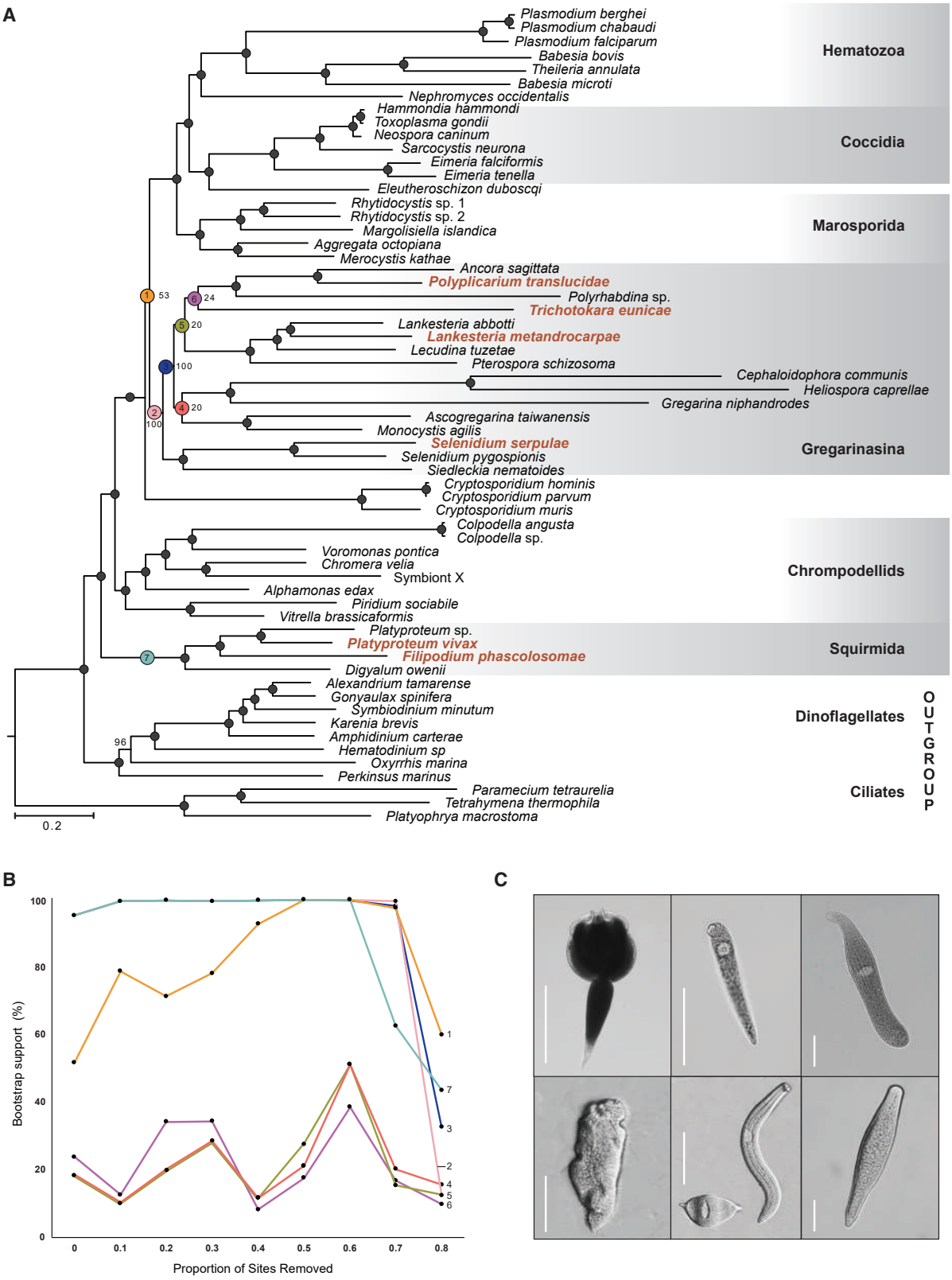

Figure 1. Maximum-likelihood phylogenetic tree of the apicomplexa and relatives

(A) Apicomplexan phylogeny based on 195 concatenated protein markers (IQ-TREE, LG+C40+F+G4 model). Species newly sequenced in this study are shown in bold with orange lettering. Circles and values at nodes correspond to non-parametric bootstrap support (1,000 replicates; IQ-TREE LG+F+R8 model). Black

(legend continued on next page)

single-cell transcriptomes from six species that represent five different families of gregarines and squirmids.

To interpret the patterns of mitochondrial evolution in the Apicomplexa, we first constructed a maximum likelihood (ML) phylogenomic tree incorporating all major apicomplexan groups, with a concatenated alignment of 195 nucleus-encoded genes, 55,369 amino acid sites, and 66 taxa (Figure 1A; Data S1). Our newly sequenced gregarine species, *Lankesteria metandrocarpae*, *Trichotokara eunicae*, *Polyplicarium translucidae*, and *Selenidium serpulae*, branched within the expected gregarine subgroups with full bootstrap support (Figures 1A and 1B).[22–25] *Platyproteum vivax* and *Filipodium phascolosomae* branched sister to *Digyalum oweni* in the squirmid clade[26] at the base of the apicomplexans and chrompodellids with full statistical support, providing further evidence that this group represents an independent origin of apicomplexan-like parasitism within the larger clade.[1,21]

Most major branches on the tree are strongly supported in both ML and Bayesian analyses (Figure S1), but the position of *Cryptosporidium* was an exception. We examined the possibility that the lack of diverse sampling within this lineage (as seen by the long branch leading to the closely related *Cryptosporidium* species in Figure 1A) may be the cause for this. We tested the effects of long branch attraction (LBA) by progressively removing the fastest evolving sites in the concatenated (195 gene) alignment in 10% increments and assessing changes to the bootstrap support (IQTREE; non-parametric bootstraps; n = 500; Figure 1C).[27] The statistical support for most strongly supported branches (e.g., the squirmids or the gregarines as a whole) and poorly supported branches (e.g., some relationships between subgroups of gregarines) were not substantially affected until most of the data were removed, as expected (Figure 1B). However, the support for the position of *Cryptosporidium* as sister to all other currently sampled apicomplexans increased significantly with the removal of fast-evolving sites (100% bootstrap support with the removal of 50%–70% of fast sites), supporting this position in the tree and suggesting that the low statistical support may be a phylogenetic artifact of LBA. Conversely, Bayesian analyses placed *Cryptosporidium* as sister to the gregarines with moderate support (PP = 0.93), with all four chains converging on this topology (Figure S1). Overall, confirmation of the exact position of *Cryptosporidium* will require further testing and ideally greater sampling of diversity within this lineage, yet the data do strongly support its exclusion from the monophyletic gregarines.

To examine the evolution of mitochondria across the apicomplexan phylogeny, hidden Markov models (HMMs) were constructed for proteins involved in aerobic respiration using curated alignments based on recent proteomics and genome sequencing studies on *Toxoplasma* and *Plasmodium*, as well as close photosynthetic relatives, *Chromera* and *Vitrella* (STAR Methods).[28–32] A combination of HMM searches, BLAST searches, and individual protein phylogenies was used to identify homologs in apicomplexan genomic and transcriptomic data (Figures 2 and 3; Data S2A). Similar to *Toxoplasma* and *Plasmodium*, the squirmids retain a complete set of TCA cycle enzymes, ETC complexes, and encode NDH2 rather than the canonical complex I (a system termed mitochondrial ETC variant a on Figure 3). In contrast, the profile of energy metabolism in gregarines is both derived and diverse. Two subgroups of gregarines, the archigregarines (*Selenidium* and *Siedleckia*) and *Monocystis*, contain a full complement of genes for aerobic respiration (ETC variant a in Figure 3). However, in all other gregarines, the functional repertoire is severely reduced; most species retain only complex II of the respiratory chain, the TCA cycle, and Fe-S biosynthesis (variant c in Figure 3). The two exceptions are *Trichotokara*, which retains the ATP synthase (complex V) and has evolved a similarly reduced respiratory chain to gastric-type *Cryptosporidium* parasites, *C. muris* and *C. andersoni* (ETC variant b in Figure 3), and *Pterospora*, *Polyrhabdina*, and *Ancora*, which have taken the reduction even further and retain only Fe-S cluster biosynthesis, similar to the mitosome of intestinal-type *Cryptosporidium* parasites, *C. parvum* and *C. hominis* (variant d in Figure 3).

We also searched for evidence of the mitochondrial pyruvate dehydrogenase (PDH) complex that converts glycolytic pyruvate to acetyl-coenzyme A (CoA) to feed the TCA cycle and the branched-chain ketoacid dehydrogenase complex (BCKDH) complex that has replaced the role of PDH in *Toxoplasma* and *Plasmodium*.[33] In the archigregarines, squirmids, and *Monocystis*, where we found canonical respiratory chains, we also retrieved complete BCKDH complexes and no evidence of the PDH complex, suggesting that they metabolize pyruvate in a manner similar to *Toxoplasma* and *Plasmodium* (Figure 2). All other taxa with reduced ETC (variants b–d) also lack PDH and either have a fragmented BCKDH complex or lack it altogether. However, we identified a pyruvate:NADP+ oxioreductase (PNO) in these taxa, which consists of a fused N-terminal pyruvate:ferredoxin oxioreductase (PFO) and a C-terminal NADPH-cytochrome P450 reductase (CPR) and is homologous to the PNO found in *Cryptosporidium*.[34] We also detected PNO in *Digyalum* and *Vitrella*, suggesting it is ancestral to the larger clade and has been lost multiple times in apicomplexans and chromerids (Figure S2). These findings suggest that gregarines with reduced mitochondrial variants b and c use PNO instead of BCKDH for the conversion of pyruvate to acetyl-CoA, which may be imported into the mitochondria for the TCA cycle, as previously proposed in *C. andersoni*.[35] Furthermore, we also detected

circles indicate 100% support; other support values are indicated on the nodes. Colored circles specify nodes that are tested for the effect of fast-evolving site removal.

(B) The effect of fast-evolving site removal on the bootstrap support of certain nodes (numbered 1–7) in the apicomplexan phylogeny. Graph shows progressive removal of fastest evolving sites in increments of 10% and the corresponding bootstrap support values (based on IQ-TREE 500 non-parametric bootstraps, LG+F+R8 model). Colored lines on the graph correspond to the specific nodes that are tested and shown on the phylogeny (above)

(C) Light micrographs of cells sequenced in this study with scale bars: top row, left to right, *Trichotokara eunicae* (240 μm), *Lankesteria metandrocarpae* (80 μm), and *Selenidium serpulae* (10 μm). Bottom row, left to right, *Filipodium phascolosomae* (30 μm), *Platyproteum vivax* (30 μm), and *Polyplicarium translucidae* (25 μm) are shown.

See also Figure S1 and Data S1.

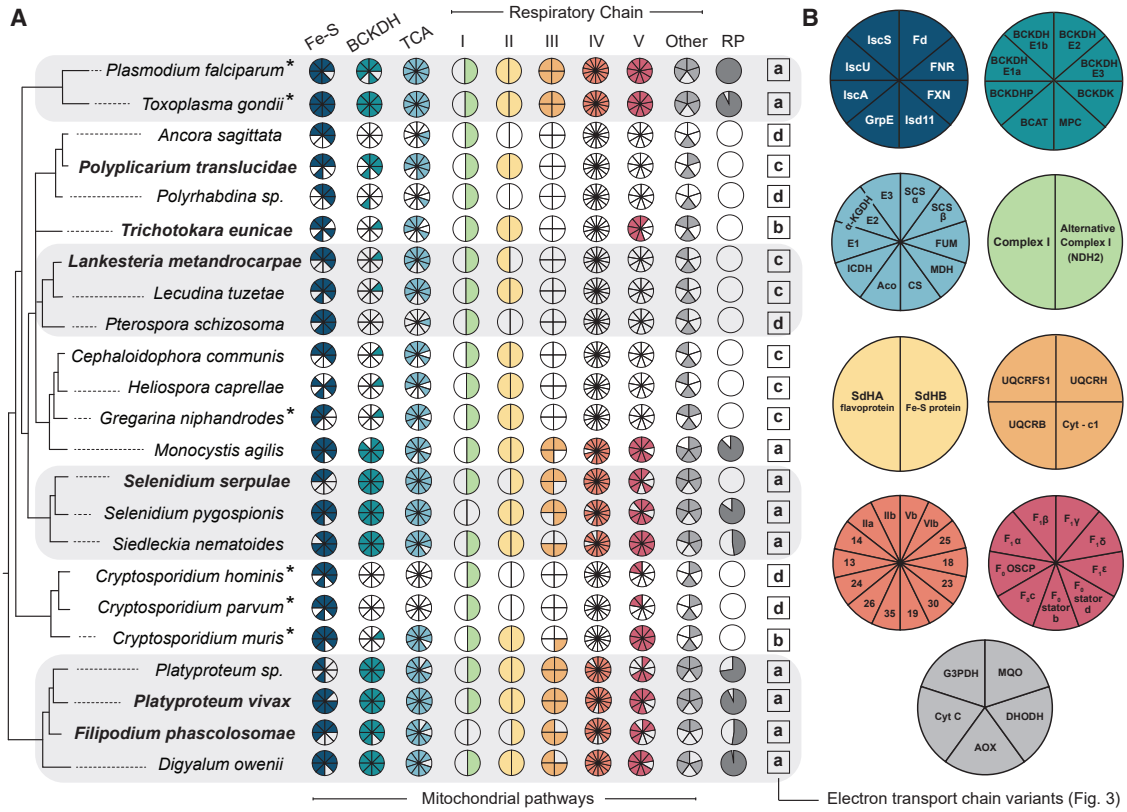

Figure 2. Distribution of mitochondrial metabolic proteins across the tree of apicomplexans and the diversity of electron transport chains

(A) Coulson plots depicting the following mitochondrial pathways: iron-sulfur cluster biosynthesis (Fe-S); branched chain ketoacid dehydrogenase (BCKDH) complex; tricarboxylic acid (TCA) cycle; respiratory chain (I-V); and other electron transport chain (ETC) proteins. Each colored pie segment represents the presence of a protein. The mitochondrial ribosomal proteins (RPs) are shown as proportions present out of the 40 proteins searched (refer to Data S2 for complete list). Species sequenced in this study are shown in orange. Asterisks (*) denote taxa with genomic sequencing data; all other taxa are represented by transcriptomic data.

(B) The proteins and subunits present in each mitochondrial complex and pathway plotted in (A). The pie chart colors correspond to the pathways in (A). Full names for all proteins and subunits can be found in Data S2.

See also Table S1, Data S2, and Figure S2.

ADP-forming acetyl-CoA synthetase (ACS) in all gregarines with reduced mitochondria (Data S2A). ACS has been widely characterized in anaerobes with MROs, where it uses acetyl-CoA to generate acetate and ATP directly.[36] However, the gregarine ACS and PNO proteins we retrieved do not contain mitochondrial localization signals, similar to those of *Cryptosporidium*.[35,37] The cellular localization of PNO in *Cryptosporidium* has been a matter of some debate, but current data suggest these proteins generate ATP by substrate-level phosphorylation in the cytosol,[37] and we predict this is also true of gregarines and that direct evidence of localization will be required to confirm this.

These data all raise the question as to whether these reduced mitochondria can possibly function in ATP generation via oxidative phosphorylation, even in taxa that contain TCA cycle enzymes and, in some cases, the ATP synthase (e.g., variant b on Figure 3). The conclusion that they cannot is bolstered by the identification of an alternative oxidase (AOX) similar to that found in *Cryptosporidium*,[11,35] the distribution of which perfectly correlates with the absence of complexes III and IV (Figure 2). AOX can accept electrons directly from coenzyme Q, which suggests the presence of a truncated AOX-dependent ETC that would not generate a proton gradient across the mitochondrial inner membrane[2] and by extension would not contribute to ATP synthesis via oxidative phosphorylation and instead might act as mechanism to cope with oxidative stress in the host intestine.[11] We also searched for *rquA*, which in some cases synthesizes the alternative quinone, rhodoquinone (RQ), and is shown in anaerobic worms to maintain a proton gradient and ATP generation in the absence of complex III.[6,38] However, no homologs of *rquA* were found in the gregarines or *Cryptosporidium*, suggesting that RQ is also not involved in ATP production in these taxa. Finally, we searched for other enzymes involved in anaerobic energy metabolism, such as [FeFe]-hydrogenase, its associated maturases, and acetate/succinate-CoA transferase (ASCT), but no evidence for these was found. Overall, most gregarines appear to lack expressed genes for enzymes that would mediate ATP synthesis in the mitochondrion.

Since they are uncultivated, much, but not all, of our current genomic data from gregarines come from single cell transcriptomes. This raises the obvious possibility that the data reflect a sampling basis rather than the absence of the genes; however,

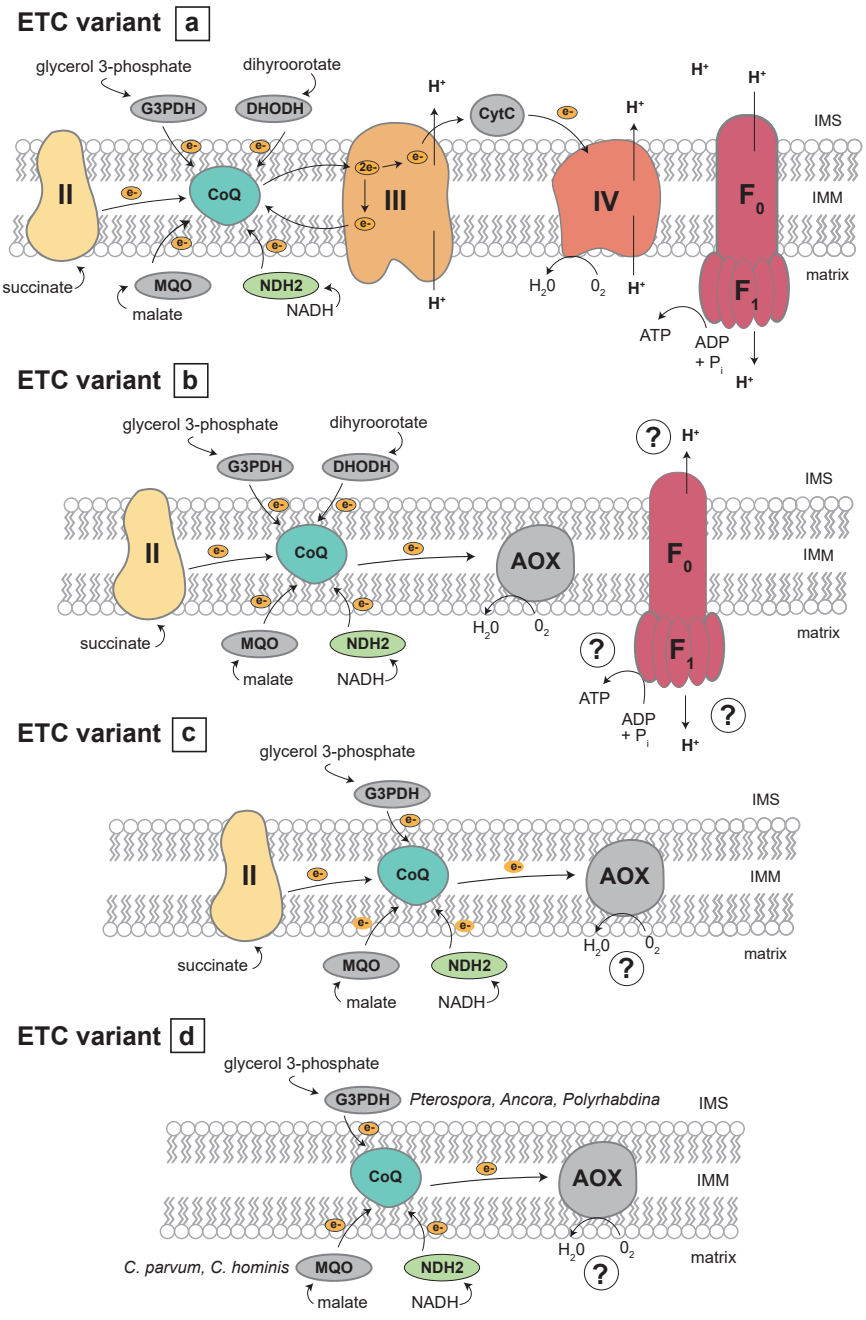

Figure 3. Schematic depicting the diversity of ETCs in apicomplexans labeled as variants a–d

The following complexes are shown in the colors corresponding to Figures 2A and 2B: complex II: succinate dehydrogenase; complex III: cytochrome *bc1* complex; complex IV: cytochrome *c* oxidase; F_1F_0-ATP synthase; AOX, alternative oxidase; as well as other dehydrogenases MQO, malate: quinone oxidoreductase; DHODH, dihydroorotate dehydrogenase; G3PDH, FAD-dependent glycerol 3-phosphate dehydrogenase; NDH2, type II NADH dehydrogenase; and mobile electron carriers CoQ, coenzyme Q; and the intermembrane space protein *cytochrome c*. The flow of electrons (shown in yellow) across the ETC complexes are depicted by arrows, and the path of protons (H^+) across the inner mitochondrial membrane (IMM) from the matrix to the intermembrane space (IMS) and back via complex V is also depicted. The uncertainty of whether complex V generates ATP in ETC variant b is depicted by "?". The AOX-dependent ETCs (variants c and d) do not lead to formation of a proton gradient and therefore are predicted not to contribute to ATP synthesis. In variant d, G3PDH is found only in *Ancora*, *Polyrhabdina*, and *Pterospora* and MQO only in *C. parvum* and *C. hominis*.
See also Table S1 and Data S2.

A more plausible alternative is that the pattern results from a systematic bias due to downregulated expression, rather than absence of the genes. Such regulation is most commonly associated with life-stage-dependent changes in ATP production, which is known in other apicomplexans, for example, *P. berghei*. However, this functional shift is associated with different stages of the parasite's complex life cycle as it moves between the different conditions prevailing in alternating hosts.[16] In contrast, the gregarine and squirmid transcriptomes are all derived from the same life history stage, the actively feeding trophozoite stage, and in all species sampled, this life stage restricted to the same environment of a single host. Mitochondrial function is therefore not simply linked to the trophozoite life stage: if it were, we would not observe such a wide range of functions in trophozoites. Similarly, there is no obvious environmental factor that correlates with the functional states we observed: nearly all the cells were isolated from animal gut environments (the single exception, *Monocystis*, is, however, very interesting because it was isolated from a seminal vesicle and its mitochondrial function is conspicuously not reduced). The factor that correlates most closely with the pattern of reduction is the phylogeny (Figure 1). This timescale is more consistent with evolutionary change rather than physiological change, but further studies to characterize the mitochondrial capabilities of different life history

several lines of evidence argue against this possibility. A stochastic sampling bias (e.g., simple incomplete sampling of expressed genes) is unlikely to lead to the observed pattern because there is a strong correlation between the presence or absence of a gene with other genes from the same functional complex: this may explain the absence of a few genes, but they should be randomly distributed, and entire complexes should not be absent if the overall sampling is relatively good. Analyses of highly conserved genes indicate most of the transcriptomes are indeed deeply sampled (Figure 2A; Data S1A), as do analyses of other systems, such as plastid-targeted genes[1,19,21] or the genes used in the phylogenomic matrix (Data S1B).

stages (sporozoites and gametes) and genome sequencing are both important and informative next steps.

However, the most interesting argument for these variants representing the actual mitochondrial function is that they are all compatible with metabolic models observed in other MRO-containing lineages.[38] Each state observed across the entire range of reduction corresponds to a viable variant ETC, mostly either functioning in ATP generation or dissipation of excess reducing potential via AOX (Figure 3). Indeed, the metabolism inferred in the most reduced gregarine mitochondria bears a striking resemblance to mitosomes in *Cryptosporidium*: complete genomes from four species in that genus reveal variable mitochondrial reduction, but the outcomes correspond to the same metabolic models as gregarines (Figure 3).

The two complexes lost most often in gregarines are complexes III and IV (which is also true of reduced mitochondria in general).[39–41] Three of the core proteins of these are Cox1, Cox3 (both in complex IV), and Cob (in complex III), which is significant because these are also the only three protein-coding genes encoded in any apicomplexan mitochondrial genome.[42] The absences of these complexes in *Cryptosporidium* mitosomes allowed for the complete loss of the mitochondrial genome,[7,43] so we sought evidence for this genome across the diversity of apicomplexans (STAR Methods). Highly represented transcripts for *cox1*, *cox3*, and *cob* were found in the squirmids, as well as in gregarines that encode other proteins in complex III and IV, but they were never observed in any gregarine lacking other complex III and IV genes (Data S2C). A mitochondria genome would require many supporting housekeeping proteins, so we also searched for nucleus-encoded proteins necessary for mitochondrial DNA expression, specifically 40 mitochondrial ribosomal proteins identified in *Plasmodium*[44] and the mitochondrial DNA-directed RNA polymerase (*POLRMT*). This revealed the same pattern of absence (Figure 2A; Data S2B), further supporting the absence of a genome. One noteworthy exception was *Selenidium serpulae*, which retains all ETC complexes but is missing evidence for any expressed mitochondrial ribosomal proteins and the *POLRMT* gene. Furthermore, we identified a *cox1* transcript from *S. serpulae* that contains a mitochondrial targeting signal (MTS), indicating the gene has been transferred to the nucleus and its protein product is targeted back to the mitochondria (Data S2D).[45] No transcripts of *cox3* and *cob* were detected. Taken together with the lack of mitochondrial housekeeping proteins, *S. serpulae* appears to have lost its mitochondrial genome, possibly reflecting a similar state as in *Amoebophrya*.[46,47] The retention of complex III and IV in *S. serpulae* suggests *cox3* and *cob* might also be present; we hypothesize one or both are either also in the nucleus or their respective complexes now function without them. Genomic sequence analysis will be required to determine the state of the *S. serpulae* mitochondria genome, but the overall pattern across apicomplexans is consistent with the absence of III and IV, resulting in the complete loss of the mitochondrial genome in most gregarines.

Functional reduction in MROs is usually accompanied by structural reduction, including the loss of hallmark features of mitochondria, such as cristae.[6,12,48] But this does not appear to apply to gregarines, where those examined at the ultrastructural level are generally found to contain conspicuous mitochondria with distinctive apicomplexan tubular cristae (refer to Table S1 for a summary of transmission electron microscopy [TEM] studies). Although there is variability in these studies, it is limited and does not clearly correspond to functional reduction: indeed, the most structural variability is between species of *Selenidium*, which all contain a complete respiratory chain (Table S1). Although in some other lineages, MROs that have lost some ETC components also retain cristae (e.g., *Blastocystis* and *Nyctotherus*),[40,49] cristae have not previously been found in mitosomes completely lacking the ETC. The MICOS complex has been implicated in cristae formation in a diversity of eukaryotes,[48] as have a cluster of recently identified apicomplexan-specific ATP synthase subunits, which form cyclical hexamers essential to shaping their tubular cristae.[31] We searched for all these proteins (STAR Methods) and detected MICOS and ATP synthase subunits in all species with functional variants a and b, but not in other gregarines (Figure 3; Data S2). This raises interesting questions about cristae formation in taxa with mitochondrial variants c and d, because at least some of these species (e.g., *Pterospora*, *Gregarina*, and *Lankesteria*) have been shown by TEM to retain tubular cristae in the same trophozoite life stage, yet we find no evidence for expression of the genes required for their formation.[50–52]

Altogether, these data suggest that the most diverse and early-branching lineages of apicomplexans, the gregarines, possess mitochondria with a wide range of functional reduction. Most gregarines appear to lack ATP generation in the mitochondrion and, at the extreme, are restricted to Fe-S biosynthesis—functionally equivalent to mitosomes. The unusual disconnect between functional and structural reduction of gregarine MROs is striking. Novel discoveries of MROs are generally associated with organisms that were previously little studied (e.g., *Mikrocytos* or *Brevimastigomonas*)[41,53] or already believed to be anaerobes (classic cases like parabasalids, diplomonads, or microsporidia).[38] The lack of mitochondria or presence of hydrogenosomes drew attention to these lineages, and as a result, our expectations for mitochondrial reduction are largely based on them.[3,5,6] Gregarines, by contrast, were formally described almost 200 years ago and have been investigated since, but because their mitochondria largely appeared unremarkable, they were not suspected to contain anaerobic MROs.

Another noteworthy feature of apicomplexan MROs is the parallel origins of a remarkably diverse collection of reduced forms. The importance of parallel evolution is emerging as a recurring theme in apicomplexan evolution, having been shown in major transitions associated with the evolution of their plastid organelles and even their dependence on the obligate parasitism of animals.[1,21] The functional reduction of the mitochondria adds another major system to this list and highlights new ways in which the most extreme outcomes of mitochondrial reduction may evolve.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.04.028.

ACKNOWLEDGMENTS

We thank Nick Irwin for assistance with sample collection and helpful discussions and Ryan Gawryluk, Vojtech Zarsky, and Mike Gray for insights into mitochondria and anaerobic metabolism. This work was supported by an Investigator Grant from the Gordon and Betty Moore Foundation awarded to P.J.K. (https://doi.org/10.37807/GBMF9201). V.M. was supported by the Li Tze Fong affiliated graduate fellowship from the University of British Columbia.

AUTHOR CONTRIBUTIONS

V.M. and P.J.K. conceived the study. K.C.W. and V.M. performed field collections and parasite isolations. V.M. carried out transcriptome sequencing, analyzed the data, and designed the figures. V.M. and P.J.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 9, 2020
Revised: February 18, 2021
Accepted: April 12, 2021
Published: May 10, 2021

REFERENCES

1. Mathur, V., Kolísko, M., Hehenberger, E., Irwin, N.A.T., Leander, B.S., Kristmundsson, Á., Freeman, M.A., and Keeling, P.J. (2019). Multiple independent origins of apicomplexan-like parasites. Curr. Biol. *29*, 2936–2941.e5.
2. Hayward, J.A., and van Dooren, G.G. (2019). Same same, but different: uncovering unique features of the mitochondrial respiratory chain of apicomplexans. Mol. Biochem. Parasitol. *232*, 111204.
3. Roger, A.J., Muñoz-Gómez, S.A., and Kamikawa, R. (2017). The origin and diversification of mitochondria. Curr. Biol. *27*, R1177–R1192.
4. Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria. Nature *326*, 332–333.
5. Gray, M.W. (2012). Mitochondrial evolution. Cold Spring Harb. Perspect. Biol. *4*, a011403.
6. Müller, M., Mentel, M., van Hellemond, J.J., Henze, K., Woehle, C., Gould, S.B., Yu, R.-Y., van der Giezen, M., Tielens, A.G.M., and Martin, W.F. (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol. Mol. Biol. Rev. *76*, 444–495.
7. Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., et al. (2004). Complete genome sequence of the apicomplexan, Cryptosporidium parvum. Science *304*, 441–445.
8. Tsaousis, A.D., Kunji, E.R.S., Goldberg, A.V., Lucocq, J.M., Hirt, R.P., and Embley, T.M. (2008). A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. Nature *453*, 553–556.
9. Tovar, J., León-Avila, G., Sánchez, L.B., Sutak, R., Tachezy, J., van der Giezen, M., Hernández, M., Müller, M., and Lucocq, J.M. (2003). Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. Nature *426*, 172–176.
10. Gill, E.E., Diaz-Triviño, S., Barberà, M.J., Silberman, J.D., Stechmann, A., Gaston, D., Tamas, I., and Roger, A.J. (2007). Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. Mol. Microbiol. *66*, 1306–1320.
11. Putignani, L., Tait, A., Smith, H.V., Horner, D., Tovar, J., Tetley, L., and Wastling, J.M. (2004). Characterization of a mitochondrion-like organelle in *Cryptosporidium parvum*. Parasitology *129*, 1–18.
12. Hampl, V., and Simpson, A.G.B. (2007). Possible mitochondria-related organelles in poorly-studied "amitochondriate" eukaryotes. In Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes, J. Tachezy, ed. (Springer Berlin Heidelberg), pp. 265–282.
13. Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S.C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L.D., Herman, E.K., et al. (2016). A eukaryote without a mitochondrial organelle. Curr. Biol. *26*, 1274–1284.
14. Krungkrai, J., Kanchanarithisak, R., Krungkrai, S.R., and Rochanakij, S. (2002). Mitochondrial NADH dehydrogenase from *Plasmodium falciparum* and *Plasmodium berghei*. Exp. Parasitol. *100*, 54–61.
15. Sheiner, L., Vaidya, A.B., and McFadden, G.I. (2013). The metabolic roles of the endosymbiotic organelles of *Toxoplasma* and *Plasmodium* spp. Curr. Opin. Microbiol. *16*, 452–458.
16. Sturm, A., Mollard, V., Cozijnsen, A., Goodman, C.D., and McFadden, G.I. (2015). Mitochondrial ATP synthase is dispensable in blood-stage *Plasmodium berghei* rodent malaria but essential in the mosquito phase. Proc. Natl. Acad. Sci. USA *112*, 10216–10223.
17. Painter, H.J., Morrisey, J.M., Mather, M.W., and Vaidya, A.B. (2007). Specific role of mitochondrial electron transport in blood-stage *Plasmodium falciparum*. Nature *446*, 88–91.
18. Tsaousis, A.D., and Keithly, J.S. (2019). The Mitochondrion-Related Organelles of *Cryptosporidium* Species (Springer), pp. 243–266.
19. Mathur, V., Kwong, W.K., Husnik, F., Irwin, N.A.T., Kristmundsson, Á., Gestal, C., Freeman, M., and Keeling, P.J. (2021). Phylogenomics identifies a new major subgroup of apicomplexans, Marosporida class. nov., with extreme apicoplast genome reduction. Genome Biol. Evol. *13*, evaa244.
20. Rueckert, S., Betts, E.L., and Tsaousis, A.D. (2019). The symbiotic spectrum: where do the gregarines fit? Trends Parasitol. *35*, 687–694.
21. Janouškovec, J., Paskerova, G.G., Miroliubova, T.S., Mikhailov, K.V., Birley, T., Aleoshin, V.V., and Simdyanov, T.G. (2019). Apicomplexan-like parasites are polyphyletic and widely but selectively dependent on cryptic plastid organelles. eLife *8*, 1–24.
22. Rueckert, S., Wakeman, K.C., and Leander, B.S. (2013). Discovery of a diverse clade of gregarine apicomplexans (Apicomplexa: Eugregarinorida) from Pacific eunicid and onuphid polychaetes, including descriptions of Paralecudina n. gen., *Trichotokara japonica* n. sp., and *T. eunicae* n. sp. J. Eukaryot. Microbiol. *60*, 121–136.
23. Leander, B.S. (2007). Molecular phylogeny and ultrastructure of *Selenidium serpulae* (Apicomplexa, Archigregarinia) from the calcareous tubeworm *Serpula vermicularis* (Annelida, Polychaeta, Sabellida). Zool. Scr. *36*, 213–227.
24. Rueckert, S., Wakeman, K.C., Jenke-Kodama, H., and Leander, B.S. (2015). Molecular systematics of marine gregarine apicomplexans from Pacific tunicates, with descriptions of five novel species of *Lankesteria*. Int. J. Syst. Evol. Microbiol. *65*, 2598–2614.
25. Wakeman, K.C., and Leander, B.S. (2013). Identity of environmental DNA sequences using descriptions of four novel marine gregarine parasites, *Polyplicarium* n. gen. (Apicomplexa), from capitellid polychaetes. Mar. Biodivers. *43*, 133–147.
26. Cavalier-Smith, T. (2014). Gregarine site-heterogeneous 18S rDNA trees, revision of gregarine higher classification, and the evolutionary diversification of Sporozoa. Eur. J. Protistol. *50*, 472–495.

27. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

28. Huet, D., Rajendran, E., van Dooren, G.G., and Lourido, S. (2018). Identification of cryptic subunits from an apicomplexan ATP synthase. eLife *7*, e38097.

29. Flegontov, P., Michálek, J., Janouškovec, J., Lai, D.H., Jirků, M., Hajdušková, E., Tomčala, A., Otto, T.D., Keeling, P.J., Pain, A., et al. (2015). Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. Mol. Biol. Evol. *32*, 1115–1131.

30. Seidi, A., Muellner-Wong, L.S., Rajendran, E., Tjhin, E.T., Dagley, L.F., Aw, V.Y.T., Faou, P., Webb, A.I., Tonkin, C.J., and van Dooren, G.G. (2018). Elucidating the mitochondrial proteome of *Toxoplasma gondii* reveals the presence of a divergent cytochrome *c* oxidase. eLife *7*, e38131.

31. Mühleip, A., Kock Flygaard, R., Ovciarikova, J., Lacombe, A., Fernandes, P., Sheiner, L., and Amunts, A. (2021). ATP synthase hexamer assemblies shape cristae of *Toxoplasma* mitochondria. Nat. Commun. *12*, 120.

32. Salunke, R., Mourier, T., Banerjee, M., Pain, A., and Shanmugam, D. (2018). Highly diverged novel subunit composition of apicomplexan F-type ATP synthase identified from *Toxoplasma gondii*. PLoS Biol. *16*, e2006128.

33. Oppenheim, R.D., Creek, D.J., Macrae, J.I., Modrzynska, K.K., Pino, P., Limenitakis, J., Polonais, V., Seeber, F., Barrett, M.P., Billker, O., et al. (2014). BCKDH: the missing link in apicomplexan mitochondrial metabolism is required for full virulence of *Toxoplasma gondii* and Plasmodium berghei. PLoS Pathog. *10*, e1004263.

34. Rotte, C., Stejskal, F., Zhu, G., Keithly, J.S., and Martin, W. (2001). Pyruvate : NADP+ oxidoreductase from the mitochondrion of Euglena gracilis and from the apicomplexan *Cryptosporidium parvum*: a biochemical relic linking pyruvate metabolism in mitochondriate and amitochondriate protists. Mol. Biol. Evol. *18*, 710–720.

35. Liu, S., Roellig, D.M., Guo, Y., Li, N., Frace, M.A., Tang, K., Zhang, L., Feng, Y., and Xiao, L. (2016). Evolution of mitosome metabolism and invasion-related proteins in *Cryptosporidium*. BMC Genomics *17*, 1006.

36. Stairs, C.W., Leger, M.M., and Roger, A.J. (2015). Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140326.

37. Tsaousis, A.D., and Keithly, J.S. (2019). The mitochondrion-related organelles of Cryptosporidium species. In Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Microbiology Monographs, *Volume 9*, J. Tachezy, ed. (Springer).

38. Gawryluk, R.M.R., and Stairs, C.W. (2021). Diversity of electron transport chains in anaerobic protists. Biochim. Biophys. Acta Bioenerg. *1862*, 148334.

39. Stechmann, A., Hamblin, K., Pérez-Brocal, V., Gaston, D., Richmond, G.S.S., van der Giezen, M., Clark, C.G., and Roger, A.J. (2008). Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr. Biol. *18*, 580–585.

40. de Graaf, R.M., Ricard, G., van Alen, T.A., Duarte, I., Dutilh, B.E., Burgtorf, C., Kuiper, J.W.P., van der Staay, G.W.M., Tielens, A.G.M., Huynen, M.A., and Hackstein, J.H. (2011). The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. Mol. Biol. Evol. *28*, 2379–2391.

41. Gawryluk, R.M.R., Kamikawa, R., Stairs, C.W., Silberman, J.D., Brown, M.W., and Roger, A.J. (2016). The earliest stages of mitochondrial adaptation to low oxygen revealed in a novel rhizarian. Curr. Biol. *26*, 2729–2738.

42. Hikosaka, K., Kita, K., and Tanabe, K. (2013). Diversity of mitochondrial genome structure in the phylum Apicomplexa. Mol. Biochem. Parasitol. *188*, 26–33.

43. Xu, P., Widmer, G., Wang, Y., Ozaki, L.S., Alves, J.M., Serrano, M.G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A.J., et al. (2004). The genome of *Cryptosporidium hominis*. Nature *431*, 1107–1112.

44. Gupta, A., Shah, P., Haider, A., Gupta, K., Siddiqi, M.I., Ralph, S.A., and Habib, S. (2014). Reduced ribosomes of the apicoplast and mitochondrion of *Plasmodium* spp. and predicted interactions with antibiotics. Open Biol. *4*, 140045.

45. Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance *2*, e201900429.

46. John, U., Lu, Y., Wohlrab, S., Groth, M., Janouškovec, J., Kohli, G.S., Mark, F.C., Bickmeyer, U., Farhat, S., Felder, M., et al. (2019). An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. Sci. Adv. *5*, eaav1110.

47. Kayal, E., and Smith, D.R. (2021). Is the dinoflagellate Amoebophrya really missing a mtDNA? Mol. Biol. Evol. msab041.

48. Muñoz-Gómez, S.A., Slamovits, C.H., Dacks, J.B., Baier, K.A., Spencer, K.D., and Wideman, J.G. (2015). Ancient homology of the mitochondrial contact site and cristae organizing system points to an endosymbiotic origin of mitochondrial cristae. Curr. Biol. *25*, 1489–1495.

49. Tsaousis, A.D., Yarlett, N., and Tan, K.S.W. (2019). The Mitochondrion-Related Organelles of Blastocystis (Springer), pp. 267–286.

50. Ciancio, A., Scippa, S., and Cammarano, M. (2001). Ultrastructure of trophozoites of the gregarine *Lankesteria ascidiae* (Apicornplexa: Eugregarinida) parasitic in the ascidian *Ciona intestinalis* (Protochordata). Eur. J. Protistol. *37*, 327–336.

51. Landers, S.C. (2002). The fine structure of the gamont of *Pterospora floridiensis* (Apicomplexa: Eugregarinida). J. Eukaryot. Microbiol. *49*, 220–226.

52. Toso, M.A., and Omoto, C.K. (2007). *Gregarina niphandrodes* may lack both a plastid genome and organelle. J. Eukaryot. Microbiol. *54*, 66–72.

53. Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., and Keeling, P.J. (2013). Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in rhizaria. Curr. Biol. *23*, 1541–1547.

54. Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

55. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

56. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: interrogation of genome assemblies. https://f1000research.com/articles/6-1287.

57. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12.

58. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.

59. Rambaut, A. (2014). FigTree v1.4.2, a graphical viewer of phylogenetic trees. http://tree.bio.ed.ac.uk/software/figtree/.

60. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.

61. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

62. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

63. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.

64. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.

65. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

66. Roure, B., Rodriguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. *7* (*Suppl 1*), S2.

67. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512.

68. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

69. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

70. Bateman, A.; UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. *47* (D1), D506–D515.

71. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

72. Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., Smirnov, A., Mylnikov, A.P., and Keeling, P.J. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc. Biol. Sci. *283*, 20152802.

73. Lewis, W.H., Lind, A.E., Sendra, K.M., Onsbring, H., Williams, T.A., Esteban, G.F., Hirt, R.P., Ettema, T.J.G., and Embley, T.M. (2020). Convergent evolution of hydrogenosomes from mitochondria by gene transfer and loss. Mol. Biol. Evol. *37*, 524–539.

74. Rotterová, J., Salomaki, E., Pánek, T., Bourland, W., Žihala, D., Táborský, P., Edgcomb, V.P., Beinart, R.A., Kolísko, M., and Čepička, I. (2020). Genomics of new ciliate lineages provides insight into the evolution of obligate anaerobiosis. Curr. Biol. *30*, 2037–2050.e6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Ampure XP beads (for Smart-Seq2)	Beckman Coulter	Cat#A63881
Betaine (for Smart-Seq2)	Sigma	Cat#61962
dNTP mix (for Smart-Seq2)	Fermentas	Cat#R0192
Ethanol 99.5% (vol/vol) (for Smart-Seq2)	Kemethyl	Cat#SN366915-06
First-strand buffer (for Smart-Seq2)	Invitrogen	Cat#18064-014
KAPA HiFi HotStart ReadyMix (2x) (for Smart-Seq2)	KAPA Biosystems	Cat#KK2601
Magnesium chloride (for Smart-Seq2)	Sigma	Cat#M8266
Phusion High-Fidelity PCR Master Mix with HF Buffer	New England Biolabs	Cat#M0531S
Recombinant ribonuclease inhibitor (for Smart-Seq2)	Invitrogen	Cat#10777019
Superscript II reverse transcriptase (for Smart-Seq2)	Invitrogen	Cat#18064-014
Triton X-100	Sigma	Cat#x100
UltraPure DNase/RNase-free distilled water (for Smart-Seq2)	Thermofisher	Cat#10977023
Critical commercial assays		
Nextera XT	Illumina	FC-131-1024
Deposited data		
Phylogenomics dataset	This paper	Mendeley Data: https://doi.org/10.17632/xkwscr7fwh.
Protein alignments & phylogenies	This paper	Mendeley Data: https://doi.org/10.17632/xkwscr7fwh.
Transcriptome sequencing raw reads	This paper	NCBI SRA PRJNA682485
Transcriptome assemblies & predicted proteins	This paper	Mendeley Data: https://doi.org/10.17632/xkwscr7fwh.
Experimental models: Organisms/strains		
Platyproteum vivax	This paper	N/A
Filipodium phascolosomae	This paper	N/A
Selenidium serpulae	This paper	N/A
Polyplicarium translucidae	This paper	N/A
Trichotokara eunicae	This paper	N/A
Lankesteria metandrocarpae	This paper	N/A
Oligonucleotides		
5′–AAGCAGTGGTATCAACGCAGAGTACT30VN-3′	Picelli et al.[54]	Oligo-dT30VN
5′–AAGCAGTGGTATCAACGCAGAGT–3′	Picelli et al.[54]	IS-PCR oligo

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
5′–AAGCAGTGGTATCAAC GCAGAGT ACATrGrG+G–3′	Picelli et al.[54]	TSO
Software and algorithms		
BLAST	Altschul et al.[55]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
BlobTools	Laetsch[56]	https://blobtools.readme.io/docs
Cutadapt	Martin[57]	https://cutadapt.readthedocs.io/en/stable/
FastTree	Price et al.[58]	http://www.microbesonline.org/fasttree/
FigTree	Rambaut[59]	https://beast.community/figtree
HMMER	Finn et al.[60]	http://hmmer.org/
IQ-Tree	Nguyen et al.[27]	http://www.iqtree.org/
MAFFT	Katoh and Standley[61]	https://mafft.cbrc.jp/alignment/software/
ModelFinder	Kalyaanamoorthy et al.[62]	http://www.iqtree.org/ModelFinder/
PEAR	Zhang et al.[63]	https://github.com/tseemann/PEAR
PhyloBayes	Lartillot et al.[64]	http://www.atgc-montpellier.fr/phylobayes/
RaxML	Stamatakis[65]	https://cme.h-its.org/exelixis/web/software/raxml/
ScaFoS	Roure et al.[66]	https://megasun.bch.umontreal.ca/Software/scafos/scafos.html
Transdecoder	Haas et al.[67]	https://github.com/TransDecoder/TransDecoder/wiki
trimAl	Capella-Gutierrez et al.[68]	http://trimal.cgenomics.org/
Trinity	Grabherr et al.[69]	https://github.com/trinityrnaseq/trinityrnaseq/wiki
TargetP	Armenteros et al.[45]	https://services.healthtech.dtu.dk/service.php?TargetP-2.0

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Varsha Mathur (varsha.mathur@botany.ubc.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Transcriptome sequencing raw reads generated in this study are available at the NCBI SRA PRJNA682485. Transcriptome assemblies, phylogenomics matrices, and individual protein alignments and phylogenies have been deposited in Mendeley Data:https://doi.org/10.17632/xkwscr7fwh.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Polyplicarium translucidae was isolated from the gut of the capitellid polychaete, *Notomastus tenuis*, that was collected at low tide at Boundary Bay, Tsawwassen, British Columbia (B.C), Canada. *Filipodium phascolosomae* and *Platyproteum vivax* were isolated from the gut of the peanut worm, *Phascolosoma agassizii,* collected from Ogden Point, Victoria, B.C., Canada, from sediments at a depth of 7-10 m. *Trichotokara eunicae* was collected from the same location from the gut of the polychaete, *Eunice valens*. *Selenidium serpulae* was isolated from the calcareous tubeworm, *Serpula vermicularis*, collected from the rocky pools of Grappler Inlet near the Bamfield Marine Sciences Centre, Vancouver Island, B.C., Canada. *Lankesteria metandrocarpae* was also collected at the same location from the tunicate, *Metandrocarpa taylori.* All specimens were collected in January 2019.

METHOD DETAILS

Transcriptome sequencing and assembly

The guts of the host animals were dissected with fine-tipped forceps under a low magnification stereomicroscopes. Hand-drawn glass pipettes were used to collect individual trophozoites under an inverted microscope. Trophozoites were rinsed at least three times in filtered seawater and stored in 2 μL of cell lysis buffer (0.2% Triton X-100 and RNase inhibitor (Invitrogen)). cDNA was synthetized from a pool of 2-5 cells, using the Smart-Seq2 protocol.[54] The cDNA concentration was quantified on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). Prior to high-throughput sequencing, 1 μL of the cDNA was used as a template for a PCR amplification of the V4 region of the 18S rRNA gene using Phusion High-Fidelity DNA Polymerase (New England Biolabs, Thermo Scientific). The PCR product were sequenced by Sanger dideoxy sequencing, and BLASTn searches were used to confirm the identity of the apicomplexan and avoid downstream host contamination.[55] Sequencing libraries were prepared using the Illumina Nextera XT protocol, and sequenced on the Illumina MiSeq (2x250bp) sequencer.

Raw sequencing reads were merged using PEAR v0.9.8.[63] The adaptor and primer sequences were trimmed using cutadapt v2.10 and the transcriptomes were assembled with Trinity v2.8.5.[57,69] The contigs were filtered for host contaminants using BlobTools, in addition to BLASTn and BLASTx searches against the NCBI nt database and the SWISS-PROT database, respectively.[55,56,70] Coding sequences were predicted using a combination of TransDecoder v5.5.0 and similarity searches against the SWISS-PROT database.[67] Assessment of the quality of the assembly and annotation of the transcriptomes was done with BUSCO v4.0.6 using the alveolate marker gene set (Data S1).[71]

Phylogenomics tree construction and analyses

Transcriptomes were searched for a set of 263 genes using BLASTp that have been previously used for apicomplexan phylogenomic analyses and that represent a wide range of eukaryotes.[1,55,72] The hits were filtered using an e-value threshold of 1e-20 and a query coverage of 50%. Single gene trees were then constructed to identify paralogs and contaminants using RAxML v8.2.12 (using the PROT-GAMMA-LG model) with support from 1,000 bootstraps.[65] Each single gene tree was manually viewed in FigTree v1.4.3 and contaminants and paralogous sequences were identified and removed.[59] The final cleaned gene-sets were filtered so that they contained only a maximum of 40% missing OTUs and then concatenated in SCaFoS v1.2.5.[66] The resulting concatenated alignment consists of 195 genes spanning 55,369 amino acid positions from 64 taxa (available at Mendeley Data: https://doi.org/10.17632/xkwscr7fwh).

Maximum likelihood (ML) phylogenies were generated from the concatenated alignment in IQ-TREE v1.6.12 using the heterogenous mixture LG+C40+F+G4 model.[27] Statistical support was assessed using 1,000 non-parametric bootstraps using the LG+F+R8 model as inferred by ModelFinder.[62] Bayesian analyses were conducted using PhyloBayes MPI v1.4 using the GTR matrix with the CAT infinite mixture model and four gamma rate categories after the removal of constant sites. Four MCMC chains were run for at least 10,000 generations (chain bipartition discrepancies: max diff = 0.140848) (Figure S1).[64] The robustness of the phylogenetic topology was evaluated using fast site removal to consider artifacts of long branch attraction. Site specific rates were calculated in the concatenated alignment using IQ-TREE v1.6.12 (-wsr option), and the fastest evolving sites were removed in 10% increments. Following fast site removal, ML phylogenies and non-parametric bootstraps (n = 500) were calculated in IQ-TREE v1.5.5 (Figure 1B).

Search and identification of mitochondrial proteins

Profile Hidden Markov models (HMMs) were used to identify proteins involved in mitochondrial metabolism in the gregarines and squirmids using curated alignments. These alignments were made using previously characterized mitochondrial proteins in *Toxoplasma gondii, Plasmodium falciparum, Cryptosporidium* spp., *Chromera velia* and *Vitrella brassicaformis*, as queries for BLASTp searches (e-value threshold of 1e-5).[2,28,30–32,44,48,60] For anerobic metabolic proteins, the PDH complex, ETC complex I, and MICOS proteins, that are not found apicomplexans or chromerids, BLASTp searches were conducted using ciliate and stramenopile queries.[40,48,73,74] The resulting hits were manually parsed to retain top hits from apicomplexans and chromerids, as well as dinoflagellates, ciliates and stramenopiles. The parsed hits were then aligned with MAFFT v. 7.212.[61]

Profile HMMs were generated using the curated alignments and HMM searches were conducted on our newly-sequenced transcriptomes in addition to all publicly available gregarine transcriptomes and genomes, using HMMER v3.1 with an e-value threshold of 1e-5.[60] All resulting hits were extracted and incorporated into the original alignments, realigned and trimmed, using MAFFT v7.222 (–auto option) and trimAl v1.2 (-gt 0.2), respectively.[61,68] These alignments were used to generate phylogenies in FastTree v2.1.3.[58]

The trees were then manually assessed in FigTree and contaminants, paralogs, and long-branching divergent sequences were identified and removed.[59] The remaining sequences were realigned and used to generate maximum likelihood phylogenies in IQ-TREE v.1.6.12.[27] Phylogenetic models were selected for each tree individually based on Bayesian Information Criteria using ModelFinder as implemented in IQ-TREE, and statistical support was assessed using 1,000 ultrafast bootstraps (Protein IDs can be found in Data S2, and trees and alignments are available at Mendeley Data: https://doi.org/10.17632/xkwscr7fwh).[62] Mitochondrial localization signals were analyzed using TargetP 2.0 (Data S2).[45]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical support for the phylogenomics tree was obtained using 1,000 non-parametric bootstraps (using the LG+F+R8 substitution model) and Bayesian posterior probabilities calculated after at least 10,000 generations (GTR+CAT+G4 model, four chains, chain bipartition discrepancies: max diff = 0.140848) (Figure 1A). Bootstrap support for clades following fast site removal was calculated from 500 non-parametric bootstraps using the LG+F+R8 substitution model (Figure 1C). 1,000 ultrafast bootstrap pseudoreplicates were used for the mitochondrial protein phylogenies.